Further developments in the construction of multivariate control charts

J. Mafodya

Statistics Department, University of Zimbabwe,
P.O. Box MP 167, Mount Pleasant, Harare, Zimbabwe.

This paper presents an alternative method of constructing Multivariate control charts. The method is based on a statistic referred to hereafter as the Studentised Hotelling's statistic. From simulated data and industrial data, this method seems to be better than the usual methods in that it clearly shows the out-of-control observations. That is, it seems to have more power than the procedures currently in use. In addition it extends the work of Nola D. Tracey *et al.* (1992) on individual observations, to the more practical situation where subgroups are used.

Keywords: Control charts, Multivariate, Hotelling statistic, F-distribution, Chi-square distribution.

Introduction

The development of Multivariate control charts has two stages, namely, the initial stage and the monitoring stage. The first stage involves collecting a sample from a production process and checking if each one of the sample observations is within or out of statistical control. Only those points in statistical control are retained and are used to obtain control limits for the Multivariate control chart. The second stage involves the use of the control limits obtained in the initial stage in checking whether each of the future observations are within or out of control, i.e, monitoring the process.

The methods widely used in Quality control to construct Multivariate control charts are those based on the Hotelling's T^2 statistics. In the initial stage, the F-approximation, the x^2 approximation and the beta distribution are used. The control charts based on the F and x^2 have a weakness of violating the basic assumptions or have little power in detecting the out-of control points. Tracy *et al.* (1992) obtained an exact (Beta) distribution for checking the status of each individual observation in the initial sample.

The aim of this paper is to provide another procedure for constructing multivariate control charts using the exact F-distribution.

Materials and Methods

The methods used in this paper are all based on Hotelling's T^2 statistic. The basic assumption underlying the procedures is that the multivariate observations are from a normal distribution with unknown mean μ and unknown covariance matrix Σ. The notation to be used in this paper is as follows:

p is the number of variables in each vector,
n is the number of observations,
m is the number of subgroups,
l is the size of each subgroup,
μ is the process mean and
Σ is the population covariance matrix.

Let $Y_1, Y_2, \ldots, Y_n = X_{11}, \ldots, X_{1l}, X_{21}, \ldots, X_{2l}, X_{m1}, \ldots, X_{ml}$ be a random sample of size $n = lm$ from $N_p(\mu,\Sigma)$ and

$$\overline{X}_i = \frac{1}{l}\sum_{j=1}^{l} \mathbf{X}_{ij}$$

$$\overline{\overline{X}} = \frac{1}{m}\sum_{i=1}^{m} \overline{\mathbf{X}}_i$$

$$\mathbf{S} = \frac{1}{m-1}\sum_{i=1}^{m}(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}})(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}})'$$

The observations now under consideration are the m subgroup means:

$$\overline{\mathbf{X}}_1,\ \overline{\mathbf{X}}_2,\ \ldots\ldots,\ \overline{\mathbf{X}}_m.$$

To construct a multivariate control chart for these observations, using the Hotelling statistic, the following statistic is considered:

$$T_i^2 = (\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}})'\,\mathbf{S}^{-1}(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}})$$

Control charts based on the approximate χ^2 distribution

If the number of sub-groups i.e. m is large so that $\overline{\overline{\mathbf{X}}}$ and $\mathbf{S}$ are approximately equal to their true values μ and $(1/l)\Sigma$ respectively, the Hotelling statistic

$$T^2 = (\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}})'\mathbf{S}^{-1}(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}})$$

has an approximate chi-square distribution with p degrees of freedom(Johnson and Wichern,1992). The approximate 100 $(1-\alpha)$ percent multivariate control limits are then given by

$$\mathbf{UCL} = \chi^2(1-\alpha/2, p)$$

$$\mathbf{LCL} = \chi^2(\frac{\alpha}{2}, p).$$

Control charts based on the approximate F-distribution

It can be shown, under the assumption of independence and normality, that $(m-1)\,\mathbf{S} \sim \mathbf{W}_p\,(m-1, \frac{1}{l}\Sigma)$. Thus, if $\overline{\mathbf{X}}_i$ is assumed to be independent of $\overline{\overline{\mathbf{X}}}$ and $\mathbf{S}$, which is true, for example, if $i = m+1, m+2, \ldots$ then it is easy to see that

$$T_i^2 = (\overline{\mathbf{X}_i} - \overline{\overline{\mathbf{X}}})'S^{-1}(\overline{\mathbf{X}_i} - \overline{\overline{\mathbf{X}}}) \sim \frac{p(m+1)(m-1)}{(m-p)}F(p, m-p)$$

(Seber 1984) and the 100 (1– α) percent control limits are then given by

$$\mathbf{UCL} = \frac{p(m+1)(m-1)}{m(m-p)} F(1-\frac{\alpha}{2}, p, m-p)$$

and

$$\mathbf{UCL} = \frac{p(m+1)(m-1)}{m(m-p)} F(\frac{\alpha}{2}, p, m-p)$$

It is important to note that the assumption that $\overline{\mathbf{X}}_i$ is independent of $\overline{\overline{\mathbf{X}}}$ and **S** and consequently that $\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}}$ is independent of , **S**, is not valid for $i = 1, 2, \ldots, m$ i.e. in the start-up stage. That is, the use of the approximate F-distribution in the start-up stage is inappropriate.

Control charts based on the exact Beta distribution

Tracy *et al.* (1992) pointed out that the exact distribution of T_i^2 under the assumption of normality is given by

$$T_i^2 = (\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}})' S^{-1} (\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}}) \sim \frac{(m-1)^2}{m} B\left(\frac{p}{2}, \frac{(m-p-1)}{2}\right)$$

That is T_i^2 follows a Beta distribution times a constant function of m. This means that the 100 (1–α) percent control limits are given by

$$\mathbf{UCL} = \frac{(m-1)^2}{m} B\left(1-\frac{\alpha}{2}, \frac{p}{2}, \frac{m-p-1}{2}\right)$$

and

$$\mathbf{UCL} = \frac{(m-1)^2}{m} B\left(\frac{\alpha}{2}, \frac{p}{2}, \frac{m-p-1}{2}\right)$$

Now for monitoring purposes, let $\overline{\mathbf{X}}_i$, $i = m+1, m+2, \ldots$ be a future observation or subgroup mean. Then from the foregoing, if the process is in control i.e. under the assumptions of independence and normality, $\overline{\mathbf{X}}_i$ is independent of both $\overline{\overline{\mathbf{X}}}$ and **S**, which implies that for the future values the exact F-distribution can be used to determine the control limits. That is, in the monitoring stage the control limits are given by

$$\mathbf{UCL} = \frac{p(m+1)(m-1)}{m(m-p)} F(1-\frac{\alpha}{2}, p, m-p)$$

and

$$\mathbf{UCL} = \frac{(m+1)(m-1)}{m(m-p)} F(\frac{\alpha}{2}, p, m-p).$$

Control chart based on Studentised F-distribution

The technique introduced here seeks to extend the method of Tracy *et al.* (1992) for individual observations to the more practical case of subgroup means, as well as to provide a more powerful way of detecting out-of-control points and to allay false alarms.

Like other methods based on the Hotelling's T^2 statistic, the studentised Hotelling statistic requires the observations to come from a multivariate normal distribution. The procedure is described below.

Let

$$\overline{\overline{\mathbf{X}}}_{-i} = \frac{1}{(m-1)} \sum_{j=1, j\neq i}^{m} \overline{\mathbf{X}}_j$$

and

$$\mathbf{S}_{-i} = \frac{1}{(m-2)} \sum_{j=1,j\neq i}^{m} (\overline{\mathbf{X}}_j - \overline{\overline{\mathbf{X}}}_{-i})(\overline{\mathbf{X}}_j - \overline{\overline{\mathbf{X}}}_{-i})'.$$

That is, $\overline{\overline{\mathbf{X}}}_{-i}$ and $\mathbf{S}_{-i}$ are calculated with the i^{th} observation deleted. It is clear that $\overline{\mathbf{X}}_i$ is independent of $\overline{\overline{\mathbf{X}}}_{-i}$ and $\mathbf{S}_{-i}$. Further, it follows from the assumptions of independence and normality that

$$\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}}_{-i} \sim N_p\left(0, \frac{m}{(m-1)l}\Sigma\right)$$

and

$$(m-2)\mathbf{S}_{-i} \sim W_p\left(m-2, \frac{1}{l}\Sigma\right).$$

Then

$$T_{-i}^2 = (\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}}_{-i})'\mathbf{S}_{-j}^{-1}(\overline{\mathbf{X}}_i - \overline{\overline{\mathbf{X}}}_{-i}) \sim \frac{m(m-2)p}{(m-1)(m-p-1)} F(p, m-p-1)$$

(Mafodya, 1997). Thus 100 (1 − α) percent control limits are given by

$$\mathbf{UCL} = \frac{m(m-2)p}{(m-1)(m-p-1)} F\left(1-\frac{\alpha}{2}, p, m-p-1\right)$$

and

$$\mathbf{UCL} = \frac{m(m-2)p}{(m-1)(m-p-1)} F\left(\frac{\alpha}{2}, p, m-p-1\right)$$

In the monitoring stage the T_i^2 is calculated using the mean and the covariance matrix of the retained observations. The control limits in the monitoring stage are

$$\mathbf{UCL} = \frac{(m+1)(m-1)p}{m(m-p)} F\left(\frac{\alpha}{2}, p, m-p\right)$$

and

$$\mathbf{UCL} = \frac{(m+1)(m-1)p}{m(m-p)} F\left(\frac{\alpha}{2}, p, m-p\right).$$

It is important to note that the control limits for the start-up stage, using this procedure are essentially the same as those used in the monitoring stage with m replaced by $(m + 1)$.

Results

Two data sets were used to demonstrate the power of each of the procedures outlined above. The first data set, data set 1 in Appendix B was extracted from the Tracey *et al.* (1992) *Journal of Quality Technology*, vol. 24, No.2 page 91. The second data set, data set 2 in Appendix B was collected from a food manufacturing industry in Zimbabwe. The four methods of setting up control limits described in the previous section namely, the Chi-square approximation, the F-approximation, the exact Beta distribution and the Studentised F-distribution are applied to each data set.

APPENDIX A

Definition 1: Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots\ldots\ldots, \boldsymbol{X}_k$ be independent and identically distributed as $N_p(0, \Sigma)$. Then the random matrix $\mathbf{W} = \mathbf{X}_j\mathbf{X}_j'$ is said to have a Wishart distribution, denoted by $\mathbf{W} \sim W_p(k, \Sigma)$.

Definition 2: Let $\mathbf{X} = (X_1, X_2, \ldots, X_p)'$ be a regular i.e. non-singular p-dimensional random vector such that $\mathbf{X} \sim \mathbf{N}_\mathrm{p}(\mathbf{0},\boldsymbol{\Sigma})$ and let $\mathbf{W} \sim W_p(k, \Sigma)$. Suppose that $\mathbf{X}$ and $\mathbf{W}$ are independent and that $\mathbf{W}$ is non-singular with probability 1. Then the Hotelling statistic is defined by

$$T^2 = \mathbf{X}'\mathbf{W}^{-1}\mathbf{X}.$$

Theorem 1: Let $\mathbf{X} \sim N_p(0,\Sigma)$ and suppose that Σ is non-singular. Then $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \sim \chi^2(\mathbf{p})$

Theorem 2: Let $\mathbf{X} \sim N_p(0,\Sigma)$ and $\mathbf{W} \sim W_p(k, \Sigma)$. Suppose that $\mathbf{X}$ is regular and that $\mathbf{W}$ is non-singular with probability 1. Suppose also that $\mathbf{X}$ and $\mathbf{W}$ are independent and $k \geq p$. Let T^2 be the Hotelling statistic. Then

$$T^2 \sim \frac{p}{k-p+1} F(p, k-p+1)$$

That is, T^2 has as an F-distribution with p and k-p+1 degrees of freedom.

APPENDIX B

Table 3: Data Set 1: Extracted from *Journal of Technology*.

Obs	*Var* 1	*Var* 2	*Var* 3
1	14.92	85.77	42.26
2	16.90	83.77	43.44
3	17.38	84.46	42.74
4	16.90	86.27	43.60
5	16.92	85.23	43.18
6	16.71	83.81	43.72
7	17.07	86.08	43.33
8	16.93	85.85	43.41
9	16.71	85.73	43.28
10	16.88	86.27	42.59
11	16.73	83.46	44.00
12	17.07	85.81	42.78
13	17.60	85.92	43.11
14	16.90	84.23	43.48

Analysis of data set 1

The following Table shows the calculated T_i^2 values and studentised T_{-i}^2 values.

Table 1: Hotelling statistics for the first data set.

Obs	T^2	T_{-i}^2
1	10.9257	123.240
2	2.04102	2.62963
3	5.58271	11.1187
4	3.86395	6.08398
5	0.03718	0.03993
6	2.25341	2.96608
7	1.43537	1.74401
8	1.20768	0.76727
9	0.67655	1.43661
10	2.16924	2.83101
11	4.17173	6.82450
12	1.40028	1.69578
13	2.33196	3.09422
14	0.91317	1.04508

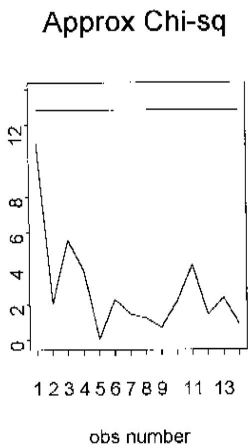

Figure 1: shows the control chart based on the approximate Chi-square distribution. The maximum value of the calculated Hotellings statistic is 10.9257.That is, all points lie below the upper control limit 12.838.

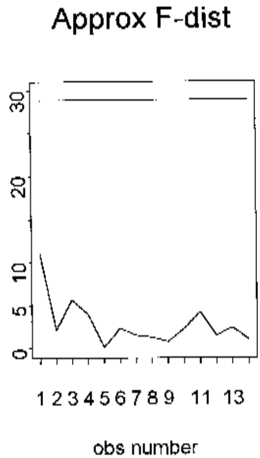

Figure 2: shows the control chart based on the approximate F-distribution. All the points are well below the upper control limit 28.872.

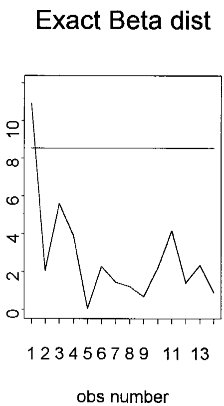

Figure 3: shows the control chart based based on the exact Beta distribution. It follows that by this procedure the first observation of this data set is out-of-control. The upper control limit is 8.456.

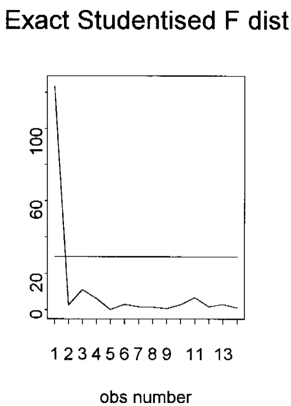

Figure 4: shows the control chart based on the Studentised F-distribution. The maximum value of the calculated studentised Hotellings statistics is 123.240 and it is very clear that the first observation is well above the upper control limit 31.328.

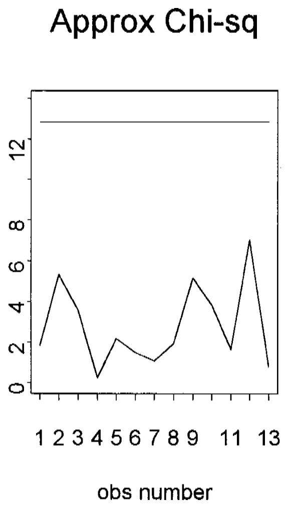

Figure 5: shows the control chart based on the approximate Chi-square distribution. All the points in the control chart are well below the upper control limit 12.838.

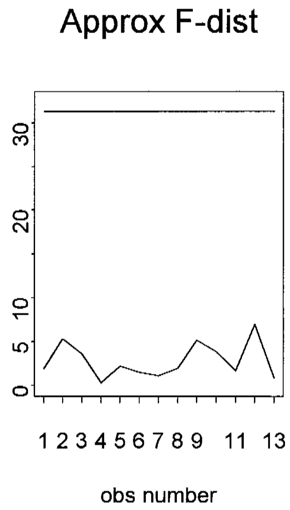

Figure 6: shows the control chart based on the approximate F-distribution. All the points are well below the upper control limit 31.328.

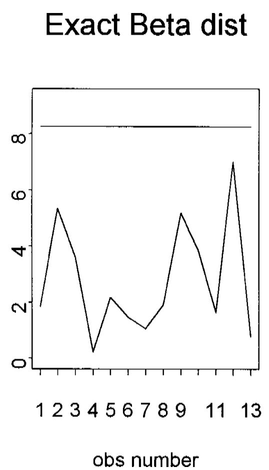

Figure 7 shows the control chart based on the Beta distribution. It can be seen from the graph that all points are well below the upper control limit 8.241.

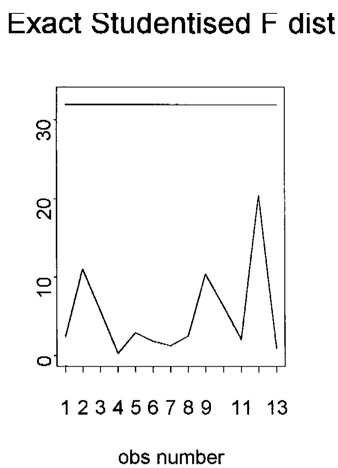

Figure 8 shows the control chart based on the Studentised F-distribution. It is clear that all points are within the control limits. The upper control limit is 31.963

A fact which may be interesting to observe from Table 1 is that all the T^2_{-i} values are greater than their corresponding T_i^2 values. The graphs on the next page shows multivariate quality control charts based on the four procedures mentioned above. Each graph shows the upper one percent upper control limit for that distribution.

That is the hypothesis of the process or this point being in control is overwhelmingly rejected. It was noted in Tracy *et al.* (1992) that this point was indeed out-of-control and that there were assignable causes for the out-of-control observation.

In the next analysis the first out of control observation is deleted from the data set. With the new data set the corresponding T_i^2 and T^2_{-i} values are calculated and are given in Table 2 below.

Table 2: Hotelling statistics for the first data set with the outlier deleted.

obs	T_i^2	T^2_{-i}
1	1.8423105	2.3773819
2	5.3295630	11.0503995
3	3.5841641	5.7003420
4	0.2316897	0.2545791
5	2.1665095	2.8974606
6	1.4635908	1.81422641
7	1.0490979	1.2467060
8	1.9143311	2.4897371
9	5.1614833	10.3978197
10	3.8377657	6.3174998
11	1.6507738	2.0869308
12	6.9981548	20.4460445
13	0.7705657	0.8909621

The graphs show the control charts based on the four methods when the first observation has been removed.

From the four control charts shown above, it is clear that all the charts are showing the same result, namely that all the observations are within the control limits.

Analysis of data set 2

The calculated values and studentised values for data set 2 are contained in Table 5 in appendix C. The graphs below show the multivariate control charts for the subgroup means.

The control charts based on the approximate chi-square distribution, Fig 9 and the control chart based on the approximate F-distribution, Fig 10 have all the points well below the upper control limit, i.e, they failed to detect the sixth point which is out of control. The control charts based on the Studentised F and beta distributions managed to detect the sixth point which is out of control.

Table 4: DATA SET 2: Food Manufacturing Industry.

Subgroup	Var1	Var2	Var3	Var4
1	520	14.5	32.7	0.08
	535	14.4	32.8	0.08
2	532	16.1	32.8	0.06
	534	15.5	32.5	0.08
3	525	15.4	32.6	0.10
	511	15.2	32.7	0.07
4	496	16.7	32.8	0.06
	524	11.6	33. 2	0.15
5	569	13.7	32.8	0.10
	475	16.4	33.4	0.19
6	506	16.0	34.0	0.22
	510	15.8	33.7	0.13
7	497	15.9	32.5	0.11
	516	15.8	32.8	0.11
8	519	15.7	32.5	0.13
	517	15.0	32.6	0.09
9	498	14.0	32.8	0.08
	499	15.7	32.7	0.13
10	502	15.5	33.0	0.09
	518	15.4	32.5	0.12
11	519	15.4	32.7	0.16
	529	14.7	32.6	0.07
12	518	14.6	32.8	0.09
	504	14.5	32.8	0.09
13	513	14.6	33.1	0.11
	502	14.5	32.6	0.13
14	506	14.8	32.8	0.08
	505	15.4	33.0	0.09
15	503	15.4	33.0	0.07
	506	15.6	32.6	0.09
16	512	15.9	32.9	0.11
	525	16.0	33.0	0.08
17	505	16.1	32.8	0.10
	508	15.8	32.9	0.09

APPENDIX C

Table 5: Hotelling Statistics for data set 2.

subgroup	T_i	T_{-i}
1	5.147	8.228
2	6.719	12.840
3	0.751	0.834
4	4.450	6.686
5	4.383	6.544
6	12.171	67.181
7	4.071	5.905
8	3.465	4.763
9	3.498	4.822
10	0.741	0.824
11	3.551	4.917
12	1.878	2.271
13	2.529	3.216
14	2.504	3.179
15	2.899	3.800
16	2.462	3.390
17	2.600	3.326

Table 6: Hotelling's statistics without the 12th subgroup.

subgroup		
1	5.215	8.802
2	6.268	12.008
3	0.969	1.105
4	4.837	7.830
5	4.382	6.761
6	9.153	27.846
7	5.057	8.386
8	3.906	5.744
9	1.155	1.337
10	3.768	5.466
11	1.569	1.997
12	2.347	2.991
13	2.426	3.113
14	2.824	3.782
15	3.035	4.111
16	2.997	4.044

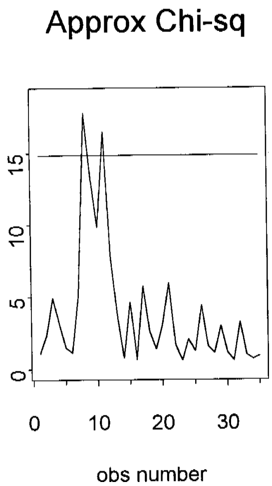

Figure 9: shows the control chart based on the approximate chi-square distribution. All the points in the control chart are well below the upper control limit 14.860.

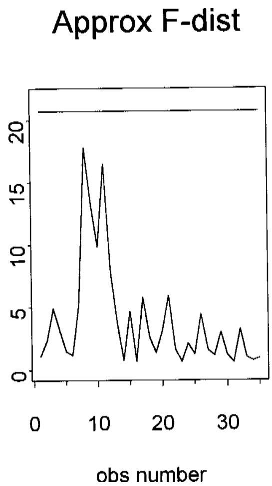

Figure 10: shows the control chart based on the approximate F-distribution. All the points are well below the upper control limit which is 32.493.

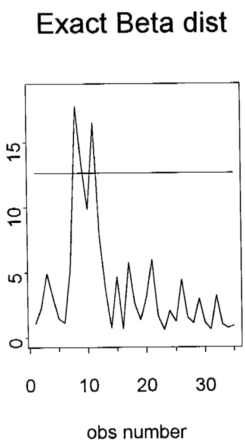

Figure 11: shows the control chart based on the Beta distribution. It can be seen from the graph that all points are below the upper control limit which is 10.314 .

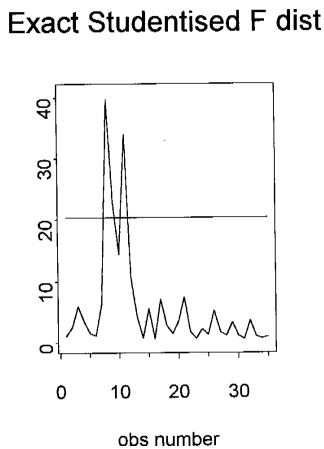

Figure 12: shows the control chart based on the Studentised F-distribution. It is clear that the sixth point is out of control. The upper control limit is 32.606

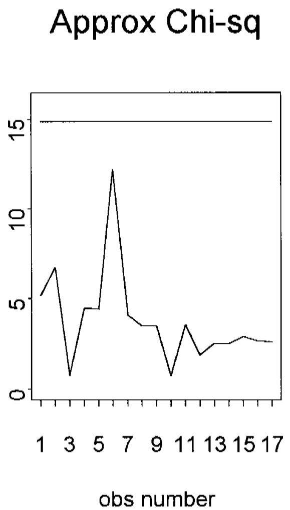

Figure 13: shows the control chart based on the approximate chi-square distribution. All the points in the control chart are well below the upper control limit 14.860. control.

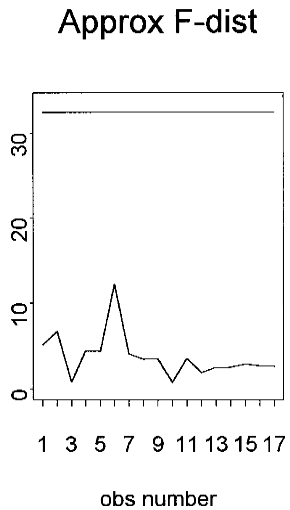

Figure 14: shows the control chart based on the approximate F-distribution. All the points are well below the upper control limit 34.644.

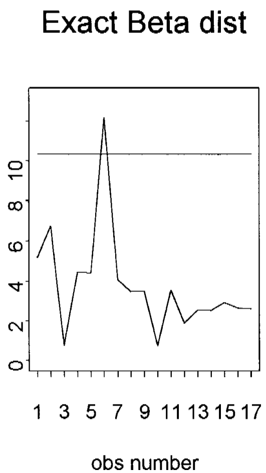

Figure 15: shows the control chart based on the Beta distribution. It can be seen from the graph that all points are well within control limits. The upper control limit is 10.047.

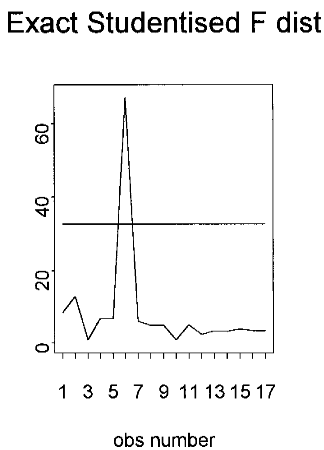

Figure 16 shows the control chart based on the Studentised F-distribution. It is clear that all points are below the upper the control limit 35.030.

In the next analysis the sixth out of control subgroup is deleted from the data set. With the new data set the corresponding T_i^2 and T_{-i}^2 values are calculated and are given in Table 6 in appendix C.

The graphs show the control charts based on the four methods when the sixth subgroup has been removed.

In all the control charts above, confirm that the process is in statistical control.

Discussion

From the multivariate control charts presented above, it was observed that the control charts based on the approximate Chi-square and the approximate F-distribution failed to detect out-of-control points for both individual points and subgroup means. The control charts based on the exact Beta distribution and exact studentised F-distribution managed to detect the out-of-control points, with the control chart based on the studentised F-distribution showing the out-of-control points more clearly than the one based on the Beta distribution.

For small samples as in data set 1, the limits based on the χ^2 distribution give misleading conclusions. It is advisable not to use the χ^2 limits in small samples. It is also clear that as the number of subgroups m increases the control limits based on the approximate F-distribution, exact Beta distribution and exact studentised distribution approach the control limits based on the χ^2 distribution.

From the various presented tables of the calculated T_i^2 and T_{-i}^2 values it is clear that for each $\overline{\mathbf{X}}_i$, the T_{-i}^2 values are greater than the corresponding T_i^2 values. The question to be addressed is, does this condition hold for all the $\overline{\mathbf{X}}_i$ observations? Several simulated data sets were examined, and in all the cases the same result was obtained namely that the T_{-i}^2 values are greater than the corresponding T_i^2 values, but the proof is thought to hinge on whether or not the difference

$$\mathbf{S_{-i}} - \mathbf{S}$$

is positive semi-definite.

Conclusion

From the results given above, the methods based on the approximate F-distribution and chi-square distributions failed to detect out-of-control situations. They should not be used for small samples and in situations where quality is of paramount importance. The control charts based on the Beta and studentised F-distributions both managed to pick up the out-of-control situations, with the control charts based on the studentised F-distribution showing the out-of-control situations more clearly than the one based on the Beta distribution.

I strongly encourage further work in comparing the effectiveness of these methods through the use power functions.

REFERENCES

MAFODYA J. 1997 Further developments in the construction of multivariate control charts.

TRACEY N.D AND YOUNG J.C. 1992, Development of multivariate control charts, Journal of Quality Technology 24 (2): page 91

JOHNSON R.A. AND WICHERN D.W. 1992. Applied Multivariate statistical analysis, 3rd ed, Prentice-Hall, International Editions.

MORRISON, D.F. 1995, Multivariate statistical methods, 3rd ed
McGraw-Hill International Editions.

Blank Page 112